

Wikibase instances in the Cultural Heritage Domain: Examples from the German humanities NFDI consortia

Introduction

The current era of Archaeology 4.0 (Thiery, 2019), also known as the *Knowledge Era* or *Era of Computing*, comprises three paradigms: (I) traditional scripts, (II) Artificial Intelligence (AI) techniques, and (III) Knowledge Graphs using technologies of I and II. To reach this era, archaeology has to go through several steps: research data stored in books (Analogue Era), applying digitisation processes to publish research data online (Digital Era), applying semantic modelling and publishing Linked Open Data (LOD) (Berners-Lee, 2006) in cooperation with community hubs (Semantic Era), and applying AI technologies including semantic reasoning to generate new knowledge (Knowledge Era) (Thiery, Veller, et al., 2023). Similar paradigms are present in other fields concerning material and immaterial Cultural Heritage (CH), including art history, architecture, and performing arts, among others (Padilla et al., 2024; Groß, 2023). To address these paradigms, the CH community requires open-source, accessible tools to manage data according to the FAIR¹ and CARE² principles.

This paper examines the use cases for adopting one specific approach and software to model community-driven data within the LOD Cloud, namely *Wikibase*. *Wikibase* is the software behind *Wikidata* (Vrandečić, 2013), the knowledge database within the Wikiverse, developed and maintained by Wikimedia Germany. It is a free and open-source software that can be used for external databases and Linked Open Data projects to share semantically structured data that both humans and machines can further reuse. We look to *Wikibase* instances developed and supported across several humanities related NFDI (German National Research Data Infrastructure) consortia to show the potential for knowledge modelling using *Wikibase* in the CH domain.

Material & Data

To create the foundation for the Semantic and Knowledge Era in the humanities, NFDI consortia (Brünger-Weilandt et al., 2020; Altenhöner et al., 2020) and Wikimedia work closely together to publish media and data as part of the Wikiverse (Fig. 1).

NFDI4Objects (Thiery, Mees, et al., 2023) uses *Wikibase* instances to model a provenance gazetteer for persons and corporate bodies³, conservation science⁴, and fuzziness and wobbliness in archaeological and geological⁵ findspots. It also contributes to *Wikidata* with the Linked Open Samian Ware⁶, the Linked Open Ogham⁷, and African Red Slip Ware digital⁸ datasets.

¹ cf. <https://www.go-fair.org/fair-principles/>

² cf. <https://www.gida-global.org/care>

³ cf. <https://n4o-prov.wikibase.cloud/wiki/Item:Q20> and <https://n4o-prov.wikibase.cloud/wiki/Item:Q23>

⁴ cf. <https://n4o-ta4-dev.wikibase.cloud/wiki/Item:Q12>

⁵ cf. <https://fuzzy-sl.wikibase.cloud/wiki/Item:Q70>

⁶ cf. https://www.wikidata.org/wiki/Wikidata:WikiProject_Linked_Open_Samian_Ware

⁷ cf. https://www.wikidata.org/wiki/Wikidata:WikiProject_Irish_Ogham_Stones

⁸ cf. https://www.wikidata.org/wiki/Wikidata:WikiProject_African_Red_Slip_Ware_Digital

NFDI4Culture provides *Wikibase* instances as a core service⁹ for the community serving a variety of use cases, ranging from backend data solution for Semantic Kompakkt¹⁰, a 3D viewing and annotation environment with a strong grounding in the art history and architecture communities (Rossenova et al., 2023), to individual research projects focusing on specific cultural collections. An NFDI4Culture satellite project develops a private *Wikibase* instance focused on documenting and preserving endangered cultural heritage in Ukraine (TIB, 2023). Additionally, the NFDI4Culture Knowledge Graph service connects with *Wikidata* for federated querying and data enrichment (Sack et al., 2023).

NFDI4Memory uses the famous *Wikibase* instance FactGrid, a database for historians, to store and provide their data.

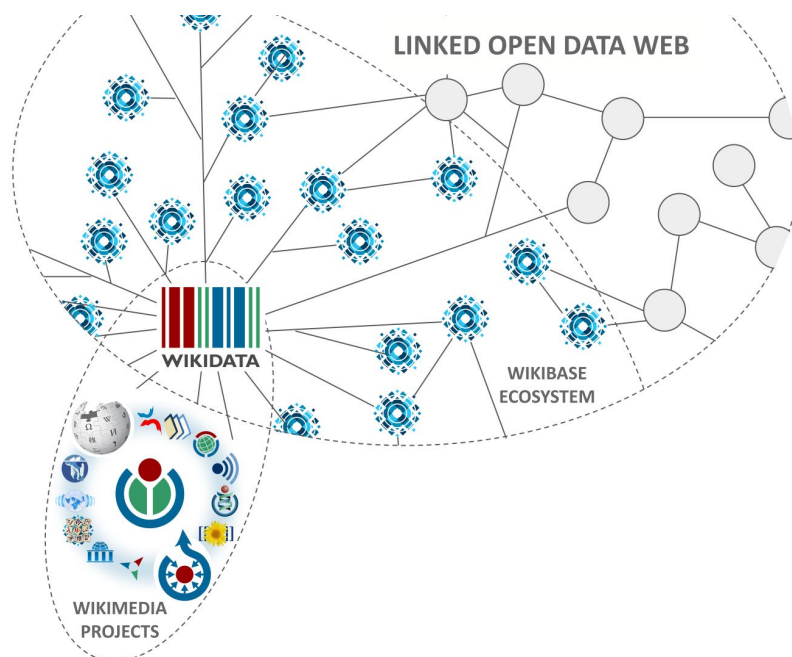


Fig. 1 – A view of the Wikimedia Linked Open Data web, including Wikiverse applications such as Wikipedia and Wikimedia Commons; Rights: Dan Shick (WMDE) CC-BY-SA 4.0.

Methodology

The above-mentioned projects adopt open science methods, including the FAIR (and CARE) data principles, through the LOD capabilities of *Wikibase*. The data model (Fig. 2) of *Wikibase* (and also *Wikidata*) structures data as semantic triples and makes these accessible to end users via a user-friendly graphical interface. It consists of entities that include items, labels or identifiers to describe them and semantic statements that attribute properties with specific values to the item. These values may be other items within the database or textual information (Bacchi and Bergamin, 2018). It is also possible to further define the primary triples with secondary statements about additional qualifications of the primary values and source references.

⁹ cf. <https://nfdi4culture.de/services/details/wikibase4research.html>

¹⁰ cf. <https://nfdi4culture.de/services/details/semantic-kompakkt.html>

This approach fits closely to the way humanities researchers are used to represent knowledge in their domains, too – as statements consisting of a subject, a predicate and an object, with the possibility to add additional argumentation to each statement (in the form of qualifiers) and to add relevant references, too – making *Wikibase* and *Wikidata* suitable environments to represent data from the CH domain (Schmidt, Thiery, and Trognitz, 2022; Rossenova, Duchesne, and Blümel, 2022).

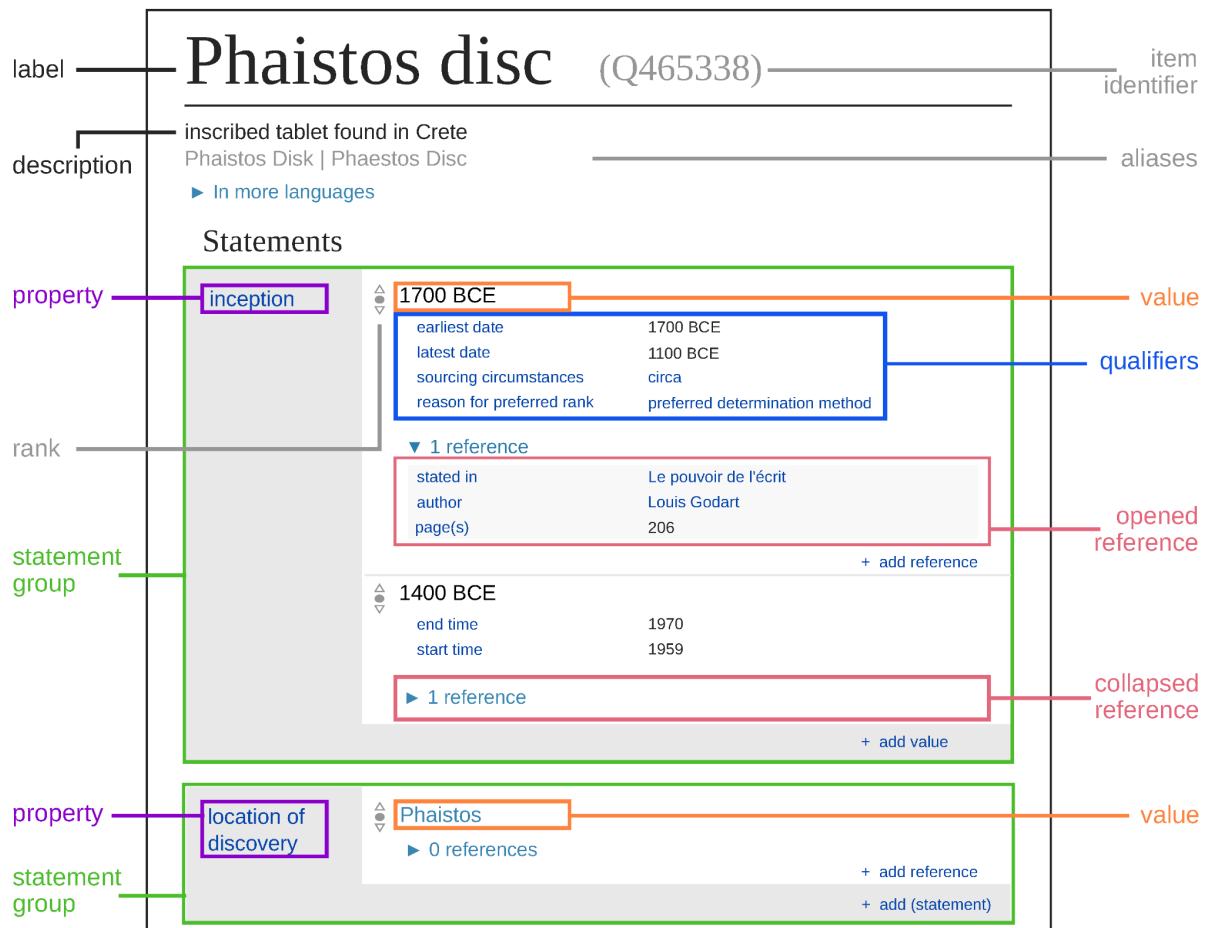


Fig. 2 – The data model in Wikidata with two statement groups and an opened reference for the item ‘Phaistos Disc’ with the identifier Q465338. Rights: Martina Trognitz, CC0, via Wikimedia Commons.

Results & Conclusions

Flexibility in how data can be structured, and accessible user interfaces provide excellent ground for wide *Wikibase* adoption in the CH domain. Still, issues around how best to harmonise and map data across this proliferating field to fully exploit the semantic capabilities for interconnection and federation remain. Theoretically, *Wikibase* instances can easily federate across each other and/or with *Wikidata* with the help of property mappings or Construct queries via the SPARQL endpoint (Rossenova, Duchesne, and Blümel, 2022). In practice, however, much of the *Wikibase* ecosystem remains relatively siloed due to the great need for more community coordination across the varied humanities disciplines and the need for ontology harmonisation through community agreements with the citizen science contributors to *Wikidata*. On the plus side, Working Groups within

individual consortia¹¹, the cross-cutting NFDI Sections¹², and the open-source developer and user communities around *Wikidata* and *Wikibase* are well aware of these challenges. They are working towards concrete solutions (Anders et al., 2022).

Discussion

Using Open Science methods, LOD and *Wikibase* software, all the material and data described above provide a wealth of semantically structured and openly accessible data ready for further reuse. The applied *Wikibase* approach helps to publish data that can meet the requirements of the Semantic and Knowledge Era paradigms and enable further AI-technology applications. The next step in achieving the full potential of this wealth of openly available CH data remains closer collaboration across individual research communities and the open-source tool communities to agree upon common, harmonised ontological standards or supervise the application of AI-supported workflows towards broader harmonisation.

References

- Altenhöner, R., I. Blümel, F. Boehm, J. Bove, K. Bicher, C. Bracht, et al. (2020), 'NFDI4Culture - Consortium for research data on material and immaterial cultural heritage', *Research Ideas and Outcomes*, 6, e57036, <https://doi.org/10.3897/rio.6.e57036>
- Anders, I., T. Arera-Rütenik, S. Arndt, R. Baum, N. Betancort, I. Blümel, et al. (2022), 'Ontology Harmonization and Mapping - Working Group Charter (NFDI section-metadata)', <https://doi.org/10.5281/ZENODO.6726519>
- Bacchi, C., and G. Bergamin (2018), 'New ways of creating and sharing bibliographic information : an experiment of using the Wikibase Data Model for UNIMARC data', *JLIS*, 3, <https://doi.org/10.4403/jlis.it-12458>
- Berners-Lee, T. (2006), 'Linked Data', <https://www.w3.org/DesignIssues/LinkedData.html> [accessed 31 May 2024]
- Brünger-Weilandt, S., K.-C. Bruhn, A. W. Busch, E. Hinrichs, G. Maier, J. Paulmann, et al. (2020), 'Memorandum of Understanding by NFDI Initiatives from the Humanities and Cultural Studies', <https://doi.org/10.5281/ZENODO.3265762>
- Gruß, M. (2023), 'Collection meets Research IV: Collection data as research data?', <https://nfdi4culture.de/id/E5204> [accessed 31 May 2024]
- Padilla, T., L. Allen, S. Varner, S. Potvin, H. Frost, and E. Russey Roke (2024), 'Always Already Computational: Collections as Data', <https://collectionsasdata.github.io/> [accessed 31 May 2024]
- Rossenova, L., P. Duchesne, and I. Blümel (2022), 'Wikidata and Wikibase as complementary research data management services for cultural heritage data', *Proceedings of the 3rd Wikidata Workshop 2022 Co-Located with the 21st International Semantic Web Conference (ISWC2022)*, <https://ceur-ws.org/Vol-3262/paper15.pdf>
- Rossenova, L., L. Sohmen, P. Duchesne, L. Günther, Z. Schubert, and I. Blümel (2023), 'Towards a common data model for semantic annotation of digital media: A new FOSS toolchain', <https://doi.org/10.5281/ZENODO.8228588>
- Sack, H., T. Schrade, O. Bruns, E. Posthumus, T. Tietz, E. Norouzi, et al. (2023), 'Knowledge Graph Based RDM Solutions: NFDI4Culture - NFDI-MatWerk - NFDI4DataScience', *Proceedings of the Conference on Research Data Infrastructure*, 1, <https://doi.org/10.52825/cordi.v1i.371>
- Schmidt, S. C., F. Thiery, and M. Trognitz (2022), 'Practices of Linked Open Data in Archaeology and Their Realisation in Wikidata', *Digital*, 2:3, 333–64, <https://doi.org/10.3390/digital2030019>

¹¹ cf. <https://nfdi4culture.de/services/details/linked-open-data-working-group.html>

¹² cf. <https://www.nfdi.de/section-metadata/>

Thiery, F. (2019), 'Archaeology 4.0: Archaeology in the Third Era of Computing', *Squirrel Papers*, 1:1, #2, <https://doi.org/10.5281/zenodo.2629595>

Thiery, F., A. W. Mees, B. Weisser, F. F. Schäfer, S. Baars, S. Nolte, et al. (2023), 'Object-Related Research Data Workflows Within NFDI4Objects and Beyond', *Proceedings of the Conference on Research Data Infrastructure*, 1, <https://doi.org/10.52825/cordi.v1i.326>

Thiery, F., J. Veller, L. Raddatz, L. Rokohl, F. Boochs, and A. W. Mees (2023), 'A Semi-Automatic Semantic-Model-Based Comparison Workflow for Archaeological Features on Roman Ceramics', *ISPRS International Journal of Geo-Information*, 12:4, 167, <https://doi.org/10.3390/ijgi12040167>

TIB (2023), 'In the media: How TIB and DDK are helping to save Ukraine's cultural heritage', <https://www.tib.eu/en/tib/news-and-events/news/details/in-the-media-how-tib-and-ddk-are-helping-to-save-ukraines-cultural-heritage> [accessed 31 May 2024]

Vrandečić, D. (2013), 'The Rise of Wikidata', *IEEE Intelligent Systems*, 28:4, 90–95, <https://doi.org/10.1109/MIS.2013.119>